DESCRIPTION

NOVEL PLANT GLYCINE AND HISTIDINE-RICH METAL-BINDING PROTEIN FAMILY AND USES THEREOF

[0001] Portions of this invention were made with government support under USDA NRICGP Grant Numbers 95-37305-2086 and 98-35305-6554. The U.S. government may have certain rights in this invention.

Cross Reference to Related Applications

[0002] This application claims the benefit of U.S. Provisional Application 60/491,939, filed August 1, 2003, which is hereby incorporated by reference in its entirety, including all figures, tables, amino acid sequences and polynucleotide sequences.

Background of the Invention

[0003] Several proteins and peptides are known to bind heavy metals and many of these are synthesized in cells as a detoxification response to the presence of specific metals or to serve transport or regulatory functions.

[0004] Pollution of ground water and soil with toxic heavy metals like mercury, cadmium, lead, etc. pose a serious health risk. Because metals are non-degradable they tend to bioaccumulate as they move up the food chain. The main sources of soil and ground water pollution are improper waste dumping, agricultural chemicals, and industrial effluents. The toxicity of these heavy metals depends on their concentration and also on their speciation. There exists a need for rapid and wide scale monitoring of heavy metals in the environment. Simple, sensitive sensors that can measure multiple elements simultaneously would be of great significance for wide scale monitoring.

Brief Summary of the Invention

[0005] The subject invention provides materials and methods for the detection and phytoremediation/bioremediation of metals from the environment. Specifically, the subject invention provides polypeptides that bind to a variety of metals as well as polynucleotides encoding such polypeptides. The subject invention provides transgenic plants containing polypeptides according to the subject invention as well as vectors and genetic constructs that

provide polypeptides, or fragments thereof, that are capable of binding various metals that can be found in the environment or at contaminated sites.

Brief Description of the Drawings

[0006] Figure 1: Exemplary peptide fragments of the invention.

[0007] Figure 2: Metal-binding constants calculated from equilibrium dialysis curves. B_{max} is measured in moles of metal/per mole monomer; average apparent K_d is presented in μM; ND means "not determined

[0008] Figures 3a-h. The vectors set forth in the table (Figure 3a) have been constructed to demonstrate the expression of AgNt84 from both inducible and constitutive promoters in such a way that its destination in the cell and *in-vivo* metal-binding properties can be determined; Gateway vectors served as the starting point for the incorporation of various genetic elements as illustrated in Figures 3b-3g and are used for demonstration purposes only. Figure 3h is an illustration of an onion epidermal cell expressing the fusion protein AgNt84-GFP from the 35S promoter of plasmid pBCAG1. The fusion protein is exported from the cell and is retained in the cell wall. The bright (lighter/whiter) areas of the cell are fluorescing GFP in the cell wall.

[0009] Figures 4a-c. The vectors set forth in the table (Figure 4a) have been constructed to demonstrate the expression of AgNt84 from both inducible and constitutive promoters in such a way that it is retained in the endoplasmic reticulum of the cell where it has the potential to be used as a metal-binding histochemical reagent with an application in electron microscopy (EM), electron energy loss spectroscopy (EELS) and secondary ion mass spectroscopy (SIMS) of biological specimens; Gateway vectors served as the starting point for the incorporation of various genetic elements as illustrated in Figures 4b. Figure 4c illustrates an onion epidermal cell expressing the fusion protein AgNt84-GFP_{HDEL} from plasmid pBCAGE1. The fusion protein is retained in the endoplasmic reticulum of the cell due to the HDEL tag. The bright (lighter/whiter) areas of the cell are fluorescing GFP in the ER.

Brief Description of the Sequences

[0010] SEQ ID NO: 1 is the polypeptide sequence of AgNt84.

[0011] SEQ ID NO: 2 is an exemplary polynucleotide sequence (cDNA) encoding AgNt84. The coding sequence of AgNt84 is located at positions 74 through 373 of SEQ ID NO: 2.

[0012] SEQ ID NO: 3 provides the amino acid sequence of the signal peptide of AgNt84.

[0013] SEQ ID NO: 4 illustrates the amino acid sequence of the metal binding domain of AgNt84.

Detailed Disclosure of the Invention

[0014] The subject invention provides isolated, recombinant, and/or purified polynucleotide sequences comprising:

a) a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NO: 1, 3, and 4;

b) a polynucleotide sequence having at least about 20% to 99.99% identity to a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NO: 1, 3, and 4, wherein said polynucleotide encodes a polypeptide having at least one of the biological activities or the polypeptides comprising SEQ ID NO: 1, 3, or 4;

c) a polynucleotide sequence encoding a biologically active fragment of a polypeptide selected from the group consisting of SEQ ID NO: 1, 3, and 4, wherein said biologically active fragment has at least one of the biological activities of the polypeptides comprising SEQ ID NO: 1, 3, or 4;

d) a polynucleotide sequence comprising SEQ ID NO: 2;

e) a polynucleotide sequence having at least about 20% to 99.99% identity to the polynucleotide sequence of SEQ ID NO: 2;

f) a polynucleotide sequence encoding a variant (*e.g.*, a variant polypeptide) of a polypeptide selected from the group consisting of SEQ ID NOs: 1, 3, and 4, wherein said variant has at least one of the biological activities associated with the polypeptides of SEQ ID NOs: 1, 3, or 4g) a polynucleotide sequence encoding a fragment of a variant polypeptide as set forth in (f);

h) a polynucleotide sequence encoding a multimeric construct;

i) a polynucleotide that is complementary to the polynucleotides set forth in (a), (b), (c), (d), (e), (f), (g), or (h);

j) a genetic construct comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), (h), or (i);

k) a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), (h), or (i);

l) a host cell comprising a vector containing a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), (h), or (i);

m) a transformed plant cell comprising a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), (h), or (i);

n) a transformed plant comprising a vector comprising a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), (h), or (i); or;

o) a polynucleotide that hybridizes under low, intermediate or high stringency with a polynucleotide sequence as set forth in (a), (b), (c), (d), (e), (f), (g), (h), or (i).

[0015] "Nucleotide sequence", "polynucleotide" or "nucleic acid" can be used interchangeably and are understood to mean, according to the present invention, either a double-stranded DNA, a single-stranded DNA or products of transcription of the said DNAs (*e.g.*, RNA molecules). It should also be understood that the present invention does not relate to genomic polynucleotide sequences in their natural environment or natural state. The nucleic acid, polynucleotide, or nucleotide sequences of the invention can be isolated, purified (or partially purified), by separation methods including, but not limited to, ion-exchange chromatography, molecular size exclusion chromatography, or by genetic engineering methods such as amplification, subtractive hybridization, cloning, subcloning or chemical synthesis, or combinations of these genetic engineering methods.

[0016] A homologous polynucleotide or polypeptide sequence, for the purposes of the present invention, encompasses a sequence having a percentage identity with the polynucleotide or polypeptide sequences, set forth herein, of between at least (or at least about) 20.00% to 99.99% (inclusive). The aforementioned range of percent identity is to be taken as including, and providing written description and support for, any fractional percentage, in intervals of 0.01%, between 20.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two nucleic acid sequences can be distributed randomly and over the entire sequence length. For example, homologous sequences can exhibit a percent identity of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent with the sequences of the instant invention. Typically, the percent identity is calculated with reference to the full length, native, and/or naturally occurring polynucleotide. The terms "identical" or percent "identity", in the context of two or more polynucleotide or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using a sequence comparison algorithm or by manual alignment and visual inspection.

[0017] Both protein and nucleic acid sequence homologies may be evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA 85(8)*:2444-2448; Altschul *et al.*, 1990, *J. Mol. Biol. 215(3)*:403-410; Thompson *et al.*, 1994, *Nucleic Acids Res. 22(2)*:4673-4680; Higgins *et al.*, 1996, *Methods Enzymol. 266*:383-402; Altschul *et al.*, 1990, *J. Mol. Biol. 215(3)*:403-410; Altschul *et al.*, 1993, *Nature Genetics 3*:266-272). Sequence comparisons are, typically, conducted using default parameters provided by the vendor or using those parameters set forth in the above-identified references, which are hereby incorporated by reference in their entireties.

[0018] A "complementary" polynucleotide sequence, as used herein, generally refers to a sequence arising from the hydrogen bonding between a particular purine and a particular pyrimidine in double-stranded nucleic acid molecules (DNA-DNA, DNA-RNA, or RNA-RNA).

The major specific pairings are guanine with cytosine and adenine with thymine or uracil. A "complementary" polynucleotide sequence may also be referred to as an "antisense" polynucleotide sequence or an "antisense" sequence.

[0019] Sequence homology and sequence identity can also be determined by hybridization studies under high stringency, intermediate stringency, and/or low stringency. Various degrees of stringency of hybridization can be employed. The more severe the conditions, the greater the complementarity that is required for duplex formation. Severity of conditions can be controlled by temperature, probe concentration, probe length, ionic strength, time, and the like. Preferably, hybridization is conducted under low, intermediate, or high stringency conditions by techniques well known in the art, as described, for example, in Keller, G.H., M.M. Manak [1987] *DNA Probes*, Stockton Press, New York, NY., pp. 169-170.

[0020] For example, hybridization of immobilized DNA on Southern blots with ^{32}P-labeled gene-specific probes can be performed by standard methods (Maniatis *et al.* [1982] *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York). In general, hybridization and subsequent washes can be carried out under intermediate to high stringency conditions that allow for detection of target sequences with homology to the exemplified polynucleotide sequence. For double-stranded DNA gene probes, hybridization can be carried out overnight at 20-25° C below the melting temperature (T_m) of the DNA hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1 mg/ml denatured blocking DNA. The melting temperature is described by the following formula (Beltz *et al.* [1983] *Methods of Enzymology*, R. Wu, L. Grossman and K. Moldave [eds.] Academic Press, New York 100:266-285).

[0021] For oligonucleotide probes, hybridization can be carried out overnight at 10-20°C below the melting temperature (T_m) of the hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1 mg/ml denatured blocking DNA. T_m for oligonucleotide probes can be determined by the following formula:

[0022] $Tm=81.5°C+16.6\ Log[Na^+]+0.41(\%G+C)-0.61(\%$ formamide)-600/length of duplex in base pairs.

[0023] Washes are typically carried out as follows:

(1) twice at room temperature for 15 minutes in 1X SSPE, 0.1% SDS (low stringency wash);

(2) once at T_m - 20°C for 15 minutes in 0.2X SSPE, 0.1% SDS (intermediate stringency wash).

[0024] For oligonucleotide probes, hybridization can be carried out overnight at 10-20°C below the melting temperature (T_m) of the hybrid in 6X SSPE, 5X Denhardt's solution, 0.1% SDS, 0.1 mg/ml denatured DNA. T_m for oligonucleotide probes can be determined by the following formula:

[0025] T_m(°C)=2(number T/A base pairs)$^+$4(number G/C base pairs) (Suggs *et al.* [1981] *ICN-UCLA Symp. Dev. Biol. Using Purified Genes*, D.D. Brown [ed.], Academic Press, New York, 23:683-693).

[0026] Washes can be carried out as follows:

(1) twice at room temperature for 15 minutes 1X SSPE, 0.1% SDS (low stringency wash);

2) once at the hybridization temperature for 15 minutes in 1X SSPE, 0.1% SDS (intermediate stringency wash).

[0027] In general, salt and/or temperature can be altered to change stringency. With a labeled DNA fragment >70 or so bases in length, the following conditions can be used:

Low:	1 or 2X SSPE, room temperature
Low:	1 or 2X SSPE, 42°C
Intermediate:	0.2X or 1X SSPE, 65°C
High:	0.1X SSPE, 65°C.

[0028] By way of another non-limiting example, procedures using conditions of high stringency can also be performed as follows: Pre-hybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in pre-hybridization mixture

containing 100 μg/ml denatured salmon sperm DNA and 5-20 x 10^6 cpm of ^{32}P-labeled probe.

Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1X SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2X SSC and 0.1% SDS, or 0.5X SSC and 0.1% SDS, or 0.1X SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety.

[0029] Another non-limiting example of procedures using conditions of intermediate stringency are as follows: Filters containing DNA are pre-hybridized, and then hybridized at a temperature of 60°C in the presence of a 5X SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2X SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of intermediate stringency which may be used are well known in the art and as cited in Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., pp. 9.47-9.57; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. are incorporated herein in their entirety.

[0030] Duplex formation and stability depend on substantial complementarity between the two strands of a hybrid and, as noted above, a certain degree of mismatch can be tolerated. Therefore, the probe sequences of the subject invention include mutations (both single and multiple), deletions, insertions of the described sequences, and combinations thereof, wherein said mutations, insertions and deletions permit formation of stable hybrids with the target polynucleotide of interest. Mutations, insertions and deletions can be produced in a given polynucleotide sequence in many ways, and these methods are known to an ordinarily skilled artisan. Other methods may become known in the future.

[0031] It is also well known in the art that restriction enzymes can be used to obtain functional fragments of the subject DNA sequences. For example, *Bal*31 exonuclease can be conveniently

used for time-controlled limited digestion of DNA (commonly referred to as "erase-a-base" procedures). See, for example, Maniatis[1982] *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York; Wei *et al.* [1983] *J. Biol. Chem.* 258:13006-13512.

[0032] The present invention further comprises fragments of the polynucleotide sequences of the instant invention. Representative fragments of the polynucleotide sequences according to the invention will be understood to mean any nucleotide fragment having at least 5 successive nucleotides, preferably at least 12 successive nucleotides, and still more preferably at least 15 or at least 20 successive nucleotides of the sequence from which it is derived. The upper limit for such fragments is the total number of nucleotides found in the full-length sequence encoding a particular polypeptide (*e.g.*, a polypeptide selected from the group consisting of SEQ ID NO: 1, 3, and 4). The term "successive" can be interchanged with the term "consecutive". In some embodiments, a polynucleotide fragment may be referred to as "a contiguous span of at least X nucleotides, wherein X is any integer value beginning with 5; the upper limit for such fragments is the total number of nucleotides found in the full-length sequence encoding a particular polypeptide (*e.g.*, a polypeptide selected from the group consisting of SEQ ID NO: 1, 3, and 4).

[0033] In some embodiments, the subject invention includes those fragments capable of hybridizing under various conditions of stringency conditions (*e.g.*, high or intermediate or low stringency) with a nucleotide sequence according to the invention; fragments that hybridize with a nucleotide sequence of the subject invention can be, optionally, labeled as set forth below.

[0034] The subject invention also provides methods for the identification of the presence of nucleic acids according to the subject invention in transformed host cells. In these varied embodiments, the invention provides for the detection of nucleic acids in a sample comprising contacting a sample with a nucleic acid (polynucleotide) of the subject invention (such as an RNA, mRNA, DNA, cDNA, or other nucleic acid). In a preferred embodiment, the polynucleotide is a probe that is, optionally, labeled and used in the detection system. Many methods for detection of nucleic acids exist and any suitable method for detection is encompassed by the instant invention. Typical assay formats utilizing nucleic acid hybridization includes, and are not limited to, 1) nuclear run-on assay, 2) slot blot assay, 3) northern blot assay (Alwine, *et al., Proc. Natl. Acad. Sci.* 74:5350), 4) magnetic particle separation, 5) nucleic acid or DNA chips, 6) reverse Northern blot assay, 7) dot blot assay, 8) *in situ* hybridization, 9) RNase protection assay (Melton, *et al., Nuc. Acids Res.* 12:7035 and as described in the 1998 catalog of

Ambion, Inc., Austin, Tex.), 10) ligase chain reaction, 11) polymerase chain reaction (PCR), 12) reverse transcriptase (RT)-PCR (Berchtold, *et al.*, *Nuc. Acids. Res.* 17:453), 13) differential display RT-PCR (DDRT-PCR) or other suitable combinations of techniques and assays. Labels suitable for use in these detection methodologies include, and are not limited to 1) radioactive labels, 2) enzyme labels, 3) chemiluminescent labels, 4) fluorescent labels, 5) magnetic labels, or other suitable labels, including those set forth below. These methodologies and labels are well known in the art and widely available to the skilled artisan. Likewise, methods of incorporating labels into the nucleic acids are also well known to the skilled artisan.

[0035] Thus, the subject invention also provides detection probes (*e.g.*, fragments of the disclosed polynucleotide sequences) for hybridization with a target sequence or the amplicon generated from the target sequence. Such a detection probe will comprise a contiguous/consecutive span of at least 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides. Labeled probes or primers are labeled with a radioactive compound or with another type of label as set forth above. Alternatively, non-labeled nucleotide sequences may be used directly as probes or primers; however, the sequences are generally labeled with a radioactive element (^{32}P, ^{35}S, ^{3}H, ^{125}I) or with a molecule such as biotin, acetylaminofluorene, digoxigenin, 5-bromo-deoxyuridine, or fluorescein to provide probes that can be used in numerous applications.

[0036] The polynucleotide sequences according to the invention may also be used in analytical systems, such as DNA chips. DNA chips and their uses are well known in the art and (see for example, U.S. Patent Nos. 5,561,071; 5,753,439; 6,214,545; *Schena et al.*, *BioEssays*, 1996, 18:427-431; *Bianchi et al.*, *Clin. Diagn. Virol.*, 1997, 8:199-208; each of which is hereby incorporated by reference in their entireties) and/or are provided by commercial vendors such as Affymetrix, Inc. (Santa Clara, CA). In addition, the nucleic acid sequences of the subject invention can be used as molecular weight markers in nucleic acid analysis procedures.

[0037] The subject invention also provides for modified nucleotide sequences. Modified nucleic acid sequences will be understood to mean any nucleotide sequence that has been modified, according to techniques well known to persons skilled in the art, and exhibiting modifications in relation to the native, naturally occurring nucleotide sequences.

[0038] The subject invention also provides genetic constructs comprising: a) a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NO: 1, 3, and 4; b) a polynucleotide sequence having at least about 20% to 99.99% identity to a polynucleotide sequence encoding a polypeptide sequence selected from the group consisting of SEQ ID NO: 1, 3, and 4, wherein said polynucleotide encodes a polypeptide having at least one of the biological activities or the polypeptides comprising SEQ ID NO: 1, 3, or 4; c) a polynucleotide sequence encoding a biologically active fragment of a polypeptide selected from the group consisting of SEQ ID NO: 1, 3, and 4, wherein said biologically active fragment has at least one of the biological activities of the polypeptides comprising SEQ ID NO: 1, 3, or 4; d) a polynucleotide sequence comprising SEQ ID NO: 2; e) a polynucleotide sequence having at least about 20% to 99.99% identity to the polynucleotide sequence of SEQ ID NO: 2; f) a polynucleotide sequence encoding variant (*e.g.*, a variant polypeptide) of a polypeptide selected from the group consisting of SEQ ID NOs: 1, 3, and 4, wherein said variant has at least on of the biological activities associated with the polypeptides of SEQ ID NOs: 1, 3, or 4; g) a polynucleotide sequence encoding a fragment of a variant polypeptide as set forth in (f); h) a polynucleotide sequence encoding multimeric construct; or i) a polynucleotide that is complementary to the polynucleotides set forth in (a), (b), (c), (d), (e), (f), (g), or (h). Genetic constructs of the subject invention can also contain additional regulatory elements such as promoters and enhancers and, optionally, selectable markers.

[0039] Also within the scope of the subject instant invention are vectors or expression cassettes containing polynucleotides encoding the polypeptides, set forth *supra*, operably linked to regulatory elements. The vectors and expression cassettes may contain additional transcriptional control sequences as well. The vectors and expression cassettes may further comprise selectable markers. The expression cassette may contain at least one additional gene, operably linked to control elements, to be co-transformed into the organism. Alternatively, the additional gene(s) and control element(s) can be provided on multiple expression cassettes. Such expression cassettes are provided with a plurality of restriction sites for insertion of the sequences of the invention to be under the transcriptional regulation of the regulatory regions. The expression cassette(s) may additionally contain selectable marker genes operably linked to control elements.

[0040] The expression cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of the invention, and a transcriptional and

translational termination region functional in either eukaryotes or prokaryotes. The transcriptional initiation region, the promoter, may be native or analogous, or foreign or heterologous, to the host cell. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. By "foreign" is intended that the transcriptional initiation region is not found in the native organism into which the transcriptional initiation region is introduced. As used herein, a chimeric gene comprises a coding sequence operably linked.

[0041] The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acid Res.* 15:9627-9639.

[0042] Where appropriate, the polynucleotides encoding the polypeptides set forth *supra* can be optimized for expression in the transformed. That is, the genes can be synthesized using species-preferred codons corresponding to the species of interest. Methods are available in the art for synthesizing for example, plant-preferred genes. See, for example, U. S. Patent Nos. 5,380,831 and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

[0043] The expression cassettes may additionally contain 5' leader sequences in the expression cassette construct. These 5' leader sequences can, optionally, be substituted for the signal peptide found in SEQ IN NO: 1 or those amino acids found 5' to the metal binding region of the polypeptide of SEQ ID NO: 1 (*e.g.*, amino acids 1-26 or amino acids 1 through 49 of SEQ ID NO: 1). Such leader sequences can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region), Elroy-Stein *et al.* (1989) *PNAS USA* 86:6126-6130; potyvirus leaders, for example, TEV leader (Tobacco Etch Virus), Allison *et al.* (1986); MDMV Leader (Maize Dwarf Mosaic Virus), *Virology* 154:9-20; human immunoglobulin heavy-chain binding protein (BiP), Macejak *et al.* (1991) *Nature* 353:90-94; untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), Jobling *et al.* (1987) *Nature* 325:622-625; tobacco mosaic

virus leader (TMV), Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256; and maize chlorotic mottle virus leader (MCMV), Lommel *et al.* (1991) *Virology* 81:382-385. See also, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968. Other methods known to enhance translation can also be utilized.

[0044] In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, base substitutions, *e.g.*, transitions and transversions, may be involved.

[0045] Also provided are transformed plant cells and transgenic plants which contain one or more genetic constructs, vectors, or expression cassettes comprising polynucleotides of the subject invention, or biologically active fragments thereof, operably linked to control elements. As used herein, the term "planta" includes algae and higher plants (including, but not limited to trees). Thus, algae, monocots, and dicots may be transformed with genetic constructs of the invention, expression cassettes, or vectors according to the invention. In certain preferred embodiments, trees are transformed with genetic constructs according to the subject invention.

[0046] Methods of transforming cells with genetic constructs, vectors, or expression cassettes comprising the novel polynucleotides of the invention are also provided. These methods comprise transforming a plant or plant cell with a polynucleotide according to the subject invention. Plants and plant cells may be transformed by electroporation, Agrobacterium transformation (including vacuum infiltration), engineered plant virus replicons, electrophoresis, microinjection, micro-projectile bombardment, vacuum infiltration of Agrobacterium, micro-LASER beam-induced perforation of cell wall, or simply by incubation with or without polyethylene glycol (PEG). Plants transformed with a genetic construct of the invention may be produced by standard techniques known in the art for the genetic manipulation of plants. DNA can be transformed into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by *Agrobacterium* exploiting its natural gene transferability (EP-A-270355; EP-A-0116718; *NAR* 12(22):87211-87215 (1984); Townsend *et al.*, U. S. Patent No. 5,563,055); particle or microprojectile bombardment (U. S. Patent No. 5,100,792; EP-A-444882; EP-A-434616;

Sanford *et al.* U. S. Patent No. 4,945,050; Tomes *et al.* (1995) "*Direct DNA Transfer Into Plant Cells via Microprojectile Bombardment*," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); and McCabe *et al.* (1988) *Biotechnology* 6:923-926); microinjection (WO 92/09696; WO 94/00583; EP 331083; EP 175966; Green *et al.* (1987) *Plant Tissue and Cell Culture*, Academic Press; Crossway *et al.* (1986) *Biotechniques* 4:320-334); electroporation (EP 290395; WO 8706614; Riggs *et al.* (1986) *Proc. Nat. Acad. Sci. USA* 83:5602-5606; D'Halluin (1992) *Plant Cell* 4:1495-1505); other forms of direct DNA uptake (DE 4005152; WO 9012096; U. S. Patent No. 4,684,611; Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722); liposome-mediated DNA uptake (*e.g.*, Freeman *et al.* (1984) *Plant Cell Physiol.* 29:1353); or the vortexing method (*e.g.*, Kindle (1990) *Proc. Nat. Acad. Sci USA* 87:1228). Physical methods for the transformation of plant cells are reviewed in Oard (1991) *Biotech. Adv.* 9:1-11. See generally, Weissinger *et al.* (1988) *Ann. Rev. Genet.* 22:421-477; Sanford *et al.* (1987) *Particulate Science and Technology* 5:27-37; Christou *et al.* (1988) *Plant Physiol.* 87:671-674; McCabe *et al.* (1988) *Bio/Technology* 6:923-926; Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182; Singh *et al.* (1998) *Theor. Appl. Genet.* 96:319324; Datta *et al.* (1990) *Biotechnology* 8:736-740; Klein *et al.* (1988) *Pro. Natl. Acad. Sci. USA* 85:4305-4309; Klein *et al.* (1988) *Biotechnology* 6:559-563; Tomes, U. S. Patent No. 5,240,855; Buising *et al.*, U. S. Patent Nos. 5,322,783 and 5,324,646; Klein *et al.* (1988) *Plant Physiol.* 91:440-444; Fromm *et al.* (1990) *Biotechnology* 8:833-839; Hooykaas-Van Slogteren *et al.* (1984) *Nature* (London) 311:763-764; Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349; De Wet *et al.* (1985) in *The Experimental Manipulation of Ovule Tissues*, ed. Chapman *et al.* (Longman, New York), pp. 197-209; Kaeppler *et al.* (1990) *Plant Cell Reports* 9:415-418 and Kaeppler *et al.* (1992) *Theor. Appl. Genet.* 84:560-566; Li *et al.* (1993) *Plant Cell Reports* 12:250-255; Christou and Ford (1995) *Annals of Botany* 75:407-413; and Osjoda *et al.* (1996) *Nature Biotechnology* 14:745-750; U.S. Patent No. 5,661,017; PCT/US00/10103 (WO 00/62601); all of which are herein incorporated by reference in their entireties.

[0047] *Agrobacterium* transformation is used by those skilled in the art to transform algae and dicotyledonous species. Substantial progress has been made towards the routine production of stable, fertile transgenic plants in almost all economically relevant monocot plants (Toriyama *et al.* (1988) *Bio/Technology* 6:1072-1074; Zhang, *et al.* (1988) *Plant Cell Rep.* 7:379-384; Zhang *et al.* (1988) *Theor. Appl. Genet.* 76:835-840; Shimamoto *et al.* (1989) *Nature* 338:274-276; Datta *et al.* (1990) *Bio/Technology* 8:736-740; Christou *et al.* (1991) *Biotechnology* 9:957-962; Peng *et al.*

(1991) International Rice Research Institute, Manila, Philippines, pp. 563-574; Cao *et al.* (1992) *Plant Cell Rep.* 11:585-591; Li *et al.* (1993) *Plant Cell Rep.* 12:250-255; Rathore *et al.* (1993) *Plant Mol. Biol.* 21:871-884; Fromm *et al.* (1990) *Bio/Technology* 8:833-839; Gordon-Kamm *et al.* (1990) *Plant Cell* 2:603-618; D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505; Walters *et al.* (1992) *Plant Mol. Biol.* 18:189-200; Koziel *et al.* (1993) *Biotechnology* 11:194-200; Vasil, I.K. (1994) *Plant Mol. Biol.* 25:925-937; Weeks *et al.* (1993) *Plant Physiol.* 102:1077-1084; Somers *et al.* (1992) *Bio/Technology* 10:1589-1594; WO 92/14828). In particular, *Agrobacterium* mediated transformation is now emerging also as a highly efficient transformation method in monocots (Hiei, *et al.* (1994) *The Plant Journal* 6:271-282). See also, Shimamoto, K. (1994) *Current Opinion in Biotechnology* 5:158-162; Vasil, *et al.* (1992) *Bio/Technology* 10:667-674; Vain, *et al.* (1995) *Biotechnology Advances* 13(4):653-671; Vasil *et al.* (1996) *Nature Biotechnology* 14:702.

[0048] Microprojectile bombardment, electroporation, and direct DNA uptake are preferred where *Agrobacterium* is inefficient or ineffective. Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, *e.g.*, bombardment with *Agrobacterium*-coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with *Agrobacterium* (EP-A-486233).

[0049] Following transformation, a plant may be regenerated, *e.g.*, from single cells, callus tissue or leaf discs, as is standard in the art. Almost any plant can be entirely regenerated from cells, tissues, and organs of the plant. Available techniques are reviewed in Vasil *et al.* (1984) in *Cell Culture and Somatic Cell Genetics of Plants, Vols. I, II, and III, Laboratory Procedures and Their Applications* (Academic press); and Weissbach *et al.* (1989) *Methods for Plant Mol. Biol.*

[0050] The transformed plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited, and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved.

[0051] The particular choice of a transformation technology will be determined by its efficiency to transform certain plant species as well as the experience and preference of the person practicing

the invention with a particular methodology of choice. It will be apparent to the skilled person that the particular choice of a transformation system to introduce nucleic acid into plant cells is not essential to or a limitation of the invention, nor is the choice of technique for plant regeneration.

[0052] Also according to the invention, there is provided a plant cell having the constructs of the invention. A further aspect of the present invention provides a method of making such a plant cell involving introduction of a vector including the construct into a plant cell. For integration of the construct into the plant genome, such introduction will be followed by recombination between the vector and the plant cell genome to introduce the sequence of nucleotides into the genome. RNA encoded by the introduced nucleic acid construct may then be transcribed in the cell and descendants thereof, including cells in plants regenerated from transformed material. A gene stably incorporated into the genome of a plant is passed from generation to generation to descendants of the plant, so such descendants should show the desired phenotype.

[0053] The present invention also provides a plant comprising a plant cell as disclosed. Transformed seeds and plant parts are also encompassed. As used herein, the expressions "cell," "cell line," and "cell culture" are used interchangeably and all such designations include progeny. Thus, the words "transformants" and "transformed cells" include the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to naturally occurring, deliberate, or inadvertent caused mutations. Mutant progeny that have the same function or biological activity as screened for in the originally transformed cell are included. Where distinct designations are intended, it will be clear from the context.

[0054] In addition to a plant, the present invention provides any clone of such a plant, seed, or hybrid descendants, and any part of any of these, such as cuttings or seed. The invention provides any plant propagule, that is any part which may be used in reproduction or propagation, sexual or asexual, including cuttings, seed, and so on. Also encompassed by the invention is a plant which is a sexually or asexually propagated off-spring, clone, or descendant of such a plant; or any part or propagule of said plant, off-spring, clone, or descendant. Plant extracts and derivatives are also provided.

[0055] To facilitate understanding of the invention, a number of terms are defined below. All publications, patents and patent applications cited herein, whether *supra* or infra, are hereby incorporated by reference in their entirety to the extent that the reference is not inconsistent with the teachings provided herein. As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the context clearly dictates otherwise and the terms "comprising", consisting essentially of" and "consisting of" can be substituted each for the other.

[0056] As is apparent to the routineer in this technology, the disclosed methods allow for the expression of a gene of interest in any plant. The invention thus relates generally to methods for the production of transgenic plants (both monocots and dicots). As used herein, the term "transgenic plants" refers to plants (algae, monocots, or dicots), comprising plant cells in which homologous or heterologous polynucleotides are expressed as the result of manipulation by the hand of man.

[0057] As is apparent to one of ordinary skill in the art, the peptides encoded by the disclosed herein may be encoded by multiple polynucleotide sequences because of the redundancy of the genetic code. It is well within the skill of a person trained in the art to create these alternative DNA sequences encoding the same, or essentially the same, amino acid sequences. These variant DNA sequences are within the scope of the subject invention. One skilled in the art would also recognize that any polynucleotide which is transcribed, processed, and translated the amino acid sequence of SEQ ID NO: 1 is within the scope of the invention.

[0058] The terms "purified" and "isolated", when referring to a polynucleotide, nucleotide, or nucleic acid, indicate a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecules but is not flanked by both of the coding or noncoding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs (*e.g.*, DNA excised with a restriction enzyme); (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a

recombinant nucleotide sequence that is part of a hybrid gene, *i.e.*, a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of (i) DNA molecules, (ii) transfected cells, and (iii) cell clones, *e.g.*, as these occur in a DNA library such as a cDNA or genomic DNA library.

[0059] The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule and thus includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications, such as those with uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as proteins (including for *e.g.*, nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (*e.g.*, acridine, psoralen, etc.), those containing chelators (*e.g.*, metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

[0060] "Control elements" include both "transcriptional control elements" and "translational control elements". "Transcriptional control elements" include "promoter", "enhancer", and "transcription termination" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription [Maniatis *et al.* [1987] *Science* 236:1237). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in plants, yeast, insect and mammalian cells and viruses (analogous control elements, *i.e.*, promoters, are also found in prokaryotes). The selection of a particular promoter and enhancer depends on what cell type is to be used to express the peptide of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types [for review see Voss *et al.* [1986] *Trends Biochem. Sci.* 11:287 and Maniatis *et al.* [1987] *supra*. Transcriptional control elements suitable for use in plants are well known in the art. "Translational control elements" include translational initiation regions and translational termination regions functional in plants.

[0061] A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. Strong promoters may be used to produce high levels of gene transcription. Alternatively, inducible promoters may be used to selectively active gene transcription when the appropriate signal is provided. Constitutive promoters may be utilized to continuously drive gene transcription. Tissue specific promoters may also be used in the practice of the invention in order to provide localized production of gene transcripts in a desired tissue. Developmental promoters may, likewise, be used to drive transcription of a gene during a particular developmental stage of the plant. Thus, a gene of interest can be combined with constitutive, tissue-specific, inducible, developmental, or other promoters for expression in plants depending upon the desired outcome.

[0062] Constitutive promoters include, for example, CaMV 35S promoter (Odell *et al.* (1985) *Nature* 313:810-812; rice actin (McElroy *et al.* (1990) *Plant Cell* 2:163-171; ubiquitin (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 and Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588); MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730); ALS promoter (U. S. Patent No. 5,659,026), and the like. Other constitutive promoters include those in U. S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. Each of the aforementioned patents and references is hereby incorporated by reference in its entirety.

[0063] A number of inducible promoters are known in the art. For example, a pathogen-inducible promoter can be utilized. Such promoters include those from pathogenesis-related proteins (PR proteins), which are induced following infection by a pathogen; *e.g.*, PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, etc. See, for example, Redolfi *et al.* (1983) *Neth. J. Plant Pathol.* 89:245-254; Uknes *et al.* (1992) *Plant Cell* 4:645-656; and Van Loon (1985) *Plant Mol. Virol.* 4:111-116; Marineau *et al.* (1987) *Plant Mol. Biol.* 9:335-342; Matton *et al.* (1989) *Molecular Plant-Microbe Interactions* 2:325-331; Somsisch *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:2427-2430; Somsisch *et al.* (1988) *Mol. Gen. Genet.* 2:93-98; and Yang (1996) *Proc. Natl. Acad. Sci. USA* 93:14972-14977. See also, Chen *et al.* (1996) *Plant J.* 10:955-966; Zhang *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:2507-2511; Warner *et al.* (1993) *Plant J.* 3:191-201; Siebertz *et al.* (1989) *Plant Cell* 1:961-968; U. S. Patent No. 5,750,386; Cordero *et al.* (1992) *Physiol. Mol. Plant Path.* 41:189-200; each of which is incorporated by reference in its entirety.

[0064] Wound-inducible promoters may be used in the genetic constructs of the invention. Such wound-inducible promoters include potato proteinase inhibitor (pin II) gene (Ryan (1990) *Ann. Rev. Phytopath.* 28:425-449; Duan *et al.* (1996) *Nature Biotechnology* 14:494-498; wun1 and wun2, U. S. Patent No. 5,428,148; win1 and win2 (Stanford *et al.* (1989) *Mol. Gen. Genet.* 215:200-208); systemin (McGurl *et al.* (1992) *Science* 225:1570-1573); WIP1 (Rohmeier *et al.* (1993) *Plant Mol. Biol.* 22:783-792; Eckelkamp *et al.* (1993) *FEBS Letters* 323:73-76); MPI gene (Corderok *et al.* (1994) *Plant J.* 6(2):141-150; and the like. These references are also incorporated by reference in their entireties.

[0065] Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. Depending upon the objective, the promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression; or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters are known in the art and include, but are not limited to, the maize In2-2 promoter, which is activated by benzene sulfonamide herbicide safeners; the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides; and the tobacco PR-1a promoter, which is activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter in Schena *et al.* (1991) *Proc. Natl. Acad. Sci USA* 88:10421-10425 and McNellis *et al.* (1998) *Plant J.* 14(2):247-257), and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-237, and U. S. Patent Nos. 5,814,618 and 5,789,156), herein incorporated by reference.

[0066] In certain preferred embodiments, heavy metal inducible promoters are contemplated for use in the subject invention. In this aspect of the invention, polynucleotides of the invention are ligated to heavy metal inducible promoters (such as those taught in Suzuki *et al.*, Plant Journal, 2002, 32:165-173; Hakkila *et al.*, J. Microbial Methods, 2003, 54:75-79; and/or Rutherford *et al.*, J. Biol. Chem., 1999, 25827-25832) and such constructs are introduced into plants, algae, or microbes as taught herein. In various embodiments, the metal inducible promoters are induced by metals, such as (but not limited to) Zn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Hg^{2+} and/or Cu^{2+}. Plants, algae, or microbes transformed with genetic constructs as discussed in this paragraph can be used in methods of bioremediation or phytoremediation as taught herein.

[0067] Tissue specific promoters can also be used in the practice of the subject invention. For example, leaf-specific promoters can similarly be used if desired, and are taught in references which include Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7):792-803; Hansen *et al.* (1997) *Mol. Gen. Genet.* 254(3):337-343; Russel *et al.* (1997) *Transgenic Res.* 6(2):157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3):1331-1341; Van Camp *et al.* (1996) *Plant Physiol.* 112(2):525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2):513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5)773-778; Lam (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco *et al.* (1993) *Plant Mol. Biol.* 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc. Natl. Acad. Sci USA*:90(20) 9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J.* 4(3):495-505. Alternatively, root-specific promoters are known and can be selected from the many available from the literature. See, for example, Hire *et al.* (1992) *Plant Mol. Biol.* 20(2):207-218 (soybean root-specific glutamine synthetase gene); Keller and Baumgartner (1991) *Plant Cell* 3(10):1051-1061 (root-specific control element in the GRP 1.8 gene of French bean); Sanger *et al.* (1990) *Plant Mol. Biol.* 14(3):433-443 (root-specific promoter of the mannopine synthase (MAS) gene of *Agrobacterium tumefaciens*) Miao *et al.* (1991) *Plant Cell* 3(1):11-22 (full-length cDNA clone encoding cytosolic glutamine synthetase (GS), which is expressed in roots and root nodules of soybean). Bogusz *et al.* (1990) *Plant Cell* 2(7):633-641 (root specific promoters from hemoglobin genes from the nitrogen-fixing nonlegume *Parasponia andersonii* and the related non-nitrogen-fixing nonlegume *Trema tomeniosa;* Leach and Aoyagi (1991) *Plant Science* (Limerick) 79(1):69-76 (rolC and rolD root-including genes of *Agrobacterium rhizogenes*); Teeri *et al.* (1989) *EMBO J.* 8(2):343-350 (octopine synthase and TR2' gene); (VfENOD-GRP3 gene promoter); Kuster *et al.* (1995) *Plant Mol. Biol.* 29(4):759-772 and Capana *et al.* (1994) *Plant Mol. Biol.* 25(4):681-691 (rolB promoter). See also U. S. Patent Nos. 5,837,876; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732; and 5,023,179.

[0068] Other tissue specific promoters can also be used in the practice of the subject invention (see, for example U.S. Patent No. 6,544,783). For example, xylem/vascular/tracheid-specific promoters, such as those disclosed in Milioni *et al.* (2002) *Plant Cell*, 14:2813-2824; Zhong *et al.* (1999) *Plant Cell*, 11:2139-2152; Ito *et al.* (2002) *Plant Cell*, 14:3201-3211; Parker *et al.* (2003) Development 130:2139-2148; Bourquin *et al.* (2002) Plant Cell 14:3073-3088 (each of which is hereby incorporated by reference in its entirety) can be used in the practice of the subject invention.

[0069] "Seed-preferred" promoters include both "seed-specific" promoters (those promoters active during seed development such as promoters of seed storage proteins) as well as "seed-germinating" promoters (those promoters active during seed germination). See Thompson *et al.* (1989) *Bioassays* 10:108, herein incorporated by reference. Such seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ10B1 (Maize 19 kDa zein); celA (cellulose synthase); gama-zein; Glob-1; bean β-phaseolin; napin; β-conglycinin; soybean lectin; cruciferin; maize 15 kDa zein; 22 kDa zein; 27 kDa zein; g-zein; waxy; shrunken 1; shrunken 2; globulin 1; etc.

[0070] "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

[0071] As used herein, the term "expression cassette" refers to a molecule comprising at least one coding sequence operably linked to a control sequence which includes all nucleotide sequences required for the transcription of cloned copies of the coding sequence and the translation of the mRNAs in an appropriate host cell. Such expression cassettes can be used to express eukaryotic genes in a variety of hosts such as bacteria, green algae, cyanobacteria, plant cells, fungal cells, yeast cells, insect cells and animal cells. Under the invention, expression cassettes can include, but are not limited to, cloning vectors, specifically designed plasmids, viruses or virus particles. The cassettes may further include an origin of replication for autonomous replication in host cells, selectable markers, various restriction sites, a potential for high copy number and strong promoters.

[0072] By "vector" is meant any genetic element, such as a plasmid, phage, transposon, cosmid, chromosome, virus etc., which is capable of replication when associated with the proper control elements and which can transfer gene sequences between cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

[0073] During the preparation of the constructs, the various fragments of DNA will often be cloned in an appropriate cloning vector, which allows for amplification of the DNA, modification of the DNA or manipulation of the DNA by joining or removing sequences, linkers, or the like. Preferably, the vectors will be capable of replication to at least a relatively high copy number in *E.*

coli. A number of vectors are readily available for cloning, including such vectors as pBR322, vectors of the pUC series, the M13 series vectors, and pBluescript vectors (Stratagene; La Jolla, Calif.).

[0074] In order to provide a means of selecting transformed plants or plant cells, the vectors for transformation will typically contain a selectable marker gene. Marker genes are expressible DNA sequences which express a polypeptide which resists a natural inhibition by, attenuates, or inactivates a selective substance. Examples of such substances include antibiotics and, in the case of plant cells, herbicides. Selectable markers for use in animal, bacterial, plant, fungal, yeast, and insect cells are well known in the art. Exemplary selectable markers include bacterial transposons Tn5 or Tn 601(903) conferring resistance to aminoglycosides (selection for Geneticin-resistance (G418R), mycophenolic acid resistance (MPAR) (utilizing *E. coli* guanosine phosphoribosyl transferase (gpt) encoding the enzyme XGPRT; selection is performed on medium containing MPA and xanthin), methotrexate resistance (MTXR), or cadmium-resistance which incorporates the mouse metallotheionein gene (as cDNA cassette) on the vector which detoxifies heavy metal ions by chelating them.

[0075] Alternatively, a marker gene may provide some visible indication of cell transformation. For example, it may cause a distinctive appearance or growth pattern relative to plants or plant cells not expressing the selectable marker gene in the presence of some substance, either as applied directly to the plant or plant cells or as present in the plant or plant cell growth media. The use of such a marker for identification of plant cells containing a plastid construct has been described (Svab *et al.* [1993] *supra*). Numerous additional promoter regions may also be used to drive expression of the selectable marker gene, including various plant promoters and bacterial promoters which have been shown to function in plants.

[0076] A number of other markers have been developed for use with plant cells, such as resistance to chloramphenicol, the aminoglycoside G418, hygromycin, or the like. Other genes which encode a product involved in chloroplast metabolism may also be used as selectable markers. For example, genes which provide resistance to plant herbicides such as glyphosate, bromoxynil or imidazolinone may find particular use. Such genes have been reported (Stalker *et al.* [1985] *J. Biol. Chem.* 260:4724-4728 (glyphosate resistant EPSP); Stalker *et al.* [1985] *J. Biol. Chem.* 263:6310-6314 (bromoxynil resistant nitrilase gene); and Sathasivan *et al.* [1990] *Nucl. Acids Res.* 18:2188 (AHAS imidazolinone resistance gene)).

[0077] Another aspect of the invention provides vectors for the cloning and/or the expression of a polynucleotide sequences taught herein in procaryotic or animal cells. The subject invention also provides for the expression of a polypeptide, peptide, derivative, or variant encoded by a polynucleotide sequence disclosed herein comprising the culture of a procaryotic or animal cell (a host cell) transformed with a polynucleotide of the subject invention under conditions that allow for the expression of a polypeptide, biologically active fragment, or multimeric construct encoded by said polynucleotide and, optionally, recovering the expressed polypeptide, peptide, derivative, or analog.

[0078] In this aspect of the invention, the polynucleotide sequences can be regulated by a second nucleic acid sequence so that the protein or peptide is expressed in a host cell transformed with the recombinant DNA molecule. For example, expression of a protein or peptide may be controlled by any promoter/enhancer element known in the art. Promoters which may be used to control expression include, but are not limited to, the CMV-IE promoter, the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, *et al.*, 1980, *Cell* 22:787-797), the herpes simplex thymidine kinase promoter (Wagner *et al.*, 1981, *Proc. Natl. Acad. Sci. U.S.A.* 78:1441-1445), the regulatory sequences of the metallothionein gene (Brinster *et al.*, 1982, *Nature* 296:39-42); prokaryotic vectors containing promoters such as the β-lactamase promoter (Villa-Kamaroff, *et al.*, 1978, *Proc. Natl. Acad. Sci. U.S.A.* 75:3727-3731), or the *tac* promoter (DeBoer, *et al.*, 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80:21-25); see also "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242:74-94; plant expression vectors comprising the nopaline synthetase promoter region (Herrera-Estrella *et al.*, 1983, Nature 303:209-213) or the cauliflower mosaic virus 35S RNA promoter (Gardner, *et al.*, 1981, *Nucl. Acids Res.* 9:2871), and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase (Herrera-Estrella *et al.*, 1984, *Nature* 310:115-120); promoter elements from yeast or fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, and/or the alkaline phosphatase promoter.

[0079] The vectors according to the invention are, for example, vectors of plasmid or viral origin. In a specific embodiment, a vector is used that comprises a promoter operably linked to a nucleic acid sequence encoding a polypeptide as disclosed herein, one or more origins of replication, and, optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene). Expression

vectors comprise regulatory sequences that control gene expression, including gene expression in a desired host cell. Exemplary vectors for the expression of the polypeptides of the invention include the pET-type plasmid vectors (Promega) or pBAD plasmid vectors (Invitrogen) or those provided in the examples below. Furthermore, the vectors according to the invention are useful for transforming host cells so as to clone or express the polynucleotide sequences of the invention.

[0080] The invention also encompasses the host cells transformed by a vector according to the invention. These cells may be obtained by introducing into host cells a nucleotide sequence inserted into a vector as defined above, and then culturing the said cells under conditions allowing the replication and/or the expression of the polynucleotide sequences of the subject invention.

[0081] The host cell may be chosen from eukaryotic or prokaryotic systems, such as for example bacterial cells, (Gram negative or Gram positive), yeast cells (for example, *Saccharomyces cereviseae* or *Pichia pastoris*), animal cells (such as Chinese hamster ovary (CHO) cells), plant cells (*e.g.*, algae), and/or insect cells using baculovirus vectors. In some embodiments, the host cells for expression of the polypeptides include, and are not limited to, those taught in U.S. Patent Nos. 6,319,691, 6,277,375, 5,643,570, or 5,565,335, each of which is incorporated by reference in its entirety, including all references cited within each respective patent. Host cells, as set forth herein, can be used in phyto/bio-remediation methods of the subject invention.

[0082] Furthermore, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered polypeptide may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.*, glycosylation, phosphorylation) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce an unglycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in mammalian cells can also to provide glycosylation of a protein.

[0083] The subject invention provides one or more isolated polypeptides comprising:

a) a polypeptide comprising SEQ ID NO: 1;

b) a variant polypeptide having at least about 20% to 99.99% identity to the polypeptide of SEQ ID NO: 1 and which has at least one of the biological activities associated with the polypeptide of SEQ ID NO: 1;

c) a fragment of a polypeptide or a variant polypeptide, wherein said fragment or variant has substantially the same biologic activity as the polypeptide of SEQ ID NO:1;

d) a signal peptide comprising the amino acid sequence MGYSKTFLLLGLAFAVVLLISSDVSA (SEQ ID NO:3) or a fragment thereof, optionally fused, in frame, to a heterologous polypeptide sequence;

e) a polypeptide comprising the amino acid sequence HGHRHVHGHGHGHVHGNGNEHGHGHHHGRGHPGH (SEQ ID NO:4), wherein said polypeptide has the ability to bind metal atoms;

f) a multimeric polypeptide construct (also referred to as a "fusion protein" or a "multimeric construct" herein) comprising: a) (SEQ ID NO: 1)$_x$ or (SEQ ID NO: 4)$_x$; b) [L-(SEQ ID NO:1)]$_x$ or [L-(SEQ ID NO:4)]$_x$, wherein L is a linker element joined to the polypeptide of SEQ ID NO:1 and x is an integer from 2 to 100;

g) a polypeptide comprising [(SEQ ID NO:3)-L]$_x$-Y, wherein x is an integer between 1 and 10, L is a linker element, preferable an amino acid linker element, and Y is a heterologous polypeptide sequence;

[0084] The term "peptide" may be used interchangeably with "oligopeptide" or "polypeptide" in the instant specification to designate a series of residues, typically L-amino acids, connected one to the other, typically by peptide bonds between the α-amino and carboxyl groups of adjacent amino acids. Linker elements can be joined to the polypeptides of the subject invention through peptide bonds or via chemical bonds (*e.g.*, heterobifunctional chemical linker elements).

[0085] The subject invention encompasses polypeptide fragments of the full-length polypeptides disclosed herein. Polypeptide fragments, according to the subject invention, usually comprise a contiguous span of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 22, 23,

24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, or 98 amino acids of SEQ ID NO:1. Another aspect of the invention provide for a fragment comprising a contiguous span of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, or 34 amino acids of SEQ ID NO: 4. Yet another aspect of the subject invention provides for a contiguous span of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 22, 23, 24, or 25 amino acids of SEQ ID NO: 3.

[0086] Fragments, as described herein, can be obtained by cleaving the polypeptides of the invention with a proteolytic enzyme (such as trypsin, chymotrypsin, or collagenase) or with a chemical reagent, such as cyanogen bromide (CNBr). Alternatively, polypeptide fragments can be generated in a highly acidic environment, for example at pH 2.5. Such polypeptide fragments may be equally well prepared by chemical synthesis or using hosts transformed with an expression vector according to the invention. The transformed host cells contain a nucleic acid, allowing the expression of these fragments, under the control of appropriate elements for regulation and/or expression of the polypeptide fragments. Various polypeptide fragments encompassed within the scope of the subject invention are provided in Figure 1.

[0087] In certain preferred embodiments, fragments of the polypeptides disclosed herein retain at least one property or activity of the full-length polypeptide from which the fragments are derived. Thus, fragments of the polypeptide of SEQ ID NO:1 have one or more of the following properties or activities: a) the ability to bind metal atoms, such as Zn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Hg^{2+} and/or Cu^{2+}; b) the ability to bind to the cell wall of plant cells; c) anti-microbial activity; d) the ability to be secreted extracellularly; or e) the ability to specifically bind antibodies that bind to the polypeptide of SEQ ID NO: 1. Fragments of the polypeptide of SEQ ID NO: 3 have one or more of the following properties or activities: a) the ability to direct the extracellular secretion of a polypeptide to which the fragment is attached; b) the ability to specifically bind to antibodies that bind to SEQ ID NOs: 1 and/or 3; or c) anti-microbial activity. Fragments of the polypeptide of SEQ ID NO: 4 have one or more of the following properties or activities: a) the ability to bind metal atoms, such as (but not limited to) Zn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Hg^{2+} and/or Cu^{2+}; or b) anti-microbial activity.

[0088] The polypeptides, and fragments thereof, may further comprise linker elements (linkers (L)) that facilitate the attachment of the fragments to other molecules, amino acids, or polypeptide sequences. The linkers can also be used to attach the polypeptides, or fragments thereof, to solid support matrices for use in affinity purification protocols (*e.g.*, for the removal of metal ions from samples). Non-limiting examples of "linkers" suitable for the practice of the invention include chemical linkers (such as those sold by Pierce, Rockford, IL), or peptides that allow for the connection combinations of polypeptides (see, for example, linkers such as those disclosed in U.S. Patent Nos. 6,121,424, 5,843,464, 5,750,352, and 5,990,275, hereby incorporated by reference in their entirety).

[0089] In other embodiments, the linker element (L) can amino acid sequences. In other embodiments, the peptide linker has one or more of the following characteristics: a) it allows for the free rotation of the polypeptides that it links (relative to each other); b) it is resistant or susceptible to digestion (cleavage) by proteases; and c) it does not interact with the polypeptides it joins together. In various embodiments, a multimeric construct according to the subject invention includes a peptide linker and the peptide linker is 5 to 60 amino acids in length. More preferably, the peptide linker is 10 to 30, amino acids in length; the peptide linker is 10 to 20 amino acids in length. In some embodiments, the peptide linker is 17 amino acids in length. Peptide linkers suitable for use in the subject invention are made up of amino acids selected from the group consisting of Gly, Ser, Asn, Thr and Ala. Preferably, the peptide linker includes a Gly-Ser element. In a preferred embodiment, the peptide linker comprises $(Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_y$ wherein y is 1, 2, 3, 4, 5, 6, 7, or 8. Other embodiments provide for a peptide linker comprising $((Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_y\text{-}Ser\text{-}Pro)$. In certain preferred embodiments, y is a value of 3, 4, or 5. In other preferred embodiment, the peptide linker comprises $(Ser\text{-}Ser\text{-}Ser\text{-}Ser\text{-}Gly)_y$ or $((Ser\text{-}Ser\text{-}Ser\text{-}Ser\text{-}Gly)_y\text{-}Ser\text{-}Pro)$, wherein y is 1, 2, 3, 4, 5, 6, 7, or 8. In certain preferred embodiments, y is a value of 3, 4, or 5. Where cleavable linker elements are desired, one or more cleavable linker sequences such as Factor Xa or enterokinase (Invitrogen, San Diego Calif.) can be used alone or in combination with the aforementioned linkers.

[0090] Multimeric constructs of the subject invention typically comprise a series of repeating biologically active elements, optionally interspersed with other elements. Thus, a "multimeric construct" according to the subject invention can provide a polypeptide comprising a series of the metal binding domains (SEQ ID NO: 4) that are, optionally, joined together by linker elements

(either chemical linker elements or amino acid linker elements). Some embodiments provide for multimeric constructs of SEQ ID NO: 1 or SEQ ID NO: 4 wherein no linker elements are provided (*e.g.*, a multimeric construct represented by the formula (SEQ ID NO: 1)$_x$ or (SEQ ID NO: 4)$_x$, wherein x is an integer value from 2 to 100). Yet other embodiments of the subject invention provide for multimeric constructs of SEQ ID NO: 1 joined to SEQ ID NO: 4, optionally via chemical or amino acid linking elements. Such multimeric constructs can be represented by the formula [(SEQ ID NO: 1-(L)$_c$)$_a$-L$_z$-(SEQ ID NO: 4-(L)$_d$)$_b$]$_x$, wherein a and b can be the same, or different, and are an integer from 1 to 50; x is an integer from 2 to 100; L is an optional linker element as discussed *supra*; c and d can be the same, or different, and are 0, 1, 2, 3, 4, 5, 6, 7, or 8; and z is 0, 1, 2, 3, 4, 5, 6, 7, or 8. The order and arrangement of SEQ ID NO: 1 and SEQ ID NO: 4 can be altered in any fashion and it is not necessary that the sequences alternate. Where a and b are values greater than 1, SEQ ID NO: 1 and/or SEQ ID NO: 4 can be joined together directly or via optional linker elements.

[0091] Yet other embodiments provide for multimeric constructs of SEQ ID NO:1 and/or SEQ ID NO:4 joined via optional linker sequences to other polypeptides that bind to metal ions or that are useful in bioremediation/phytoremediation processes. Such multimeric constructs can be represented by the formula [(SEQ ID NO: 1-(L)$_c$)$_a$-L$_z$-(SEQ ID NO: 4-(L)$_d$)$_b$-((PBPP)$_e$-(L)$_f$)$_g$]$_x$, wherein a and b can be the same, or different, and are an integer from 1 to 50; PBPP is another polypeptide useful in bioremediation/phytoremediation processes and e and g can be the same or different and are a value from 0 to 50 (preferably 1 to 50 and where e equals 1, g is at least 1); x is an integer from 2 to 100; L is an optional linker element as discussed *supra*; c, d, and f can be the same, or different, and are 0, 1, 2, 3, 4, 5, 6, 7, or 8; and z is 0, 1, 2, 3, 4, 5, 6, 7, or 8. The order and arrangement of PBPP, SEQ ID NO: 1, and SEQ ID NO: 4 can be altered in any fashion and it is not necessary that the sequences alternate. Where a, b and g are values greater than or equal to 1, PBPP, SEQ ID NO: 1 and/or SEQ ID NO: 4 can be joined together directly (c=d=f=0) or via optional linker elements.

[0092] A "variant" polypeptide (or polypeptide variant) is to be understood to designate polypeptides exhibiting, in relation to the natural polypeptide, certain modifications. These modifications can include a deletion, addition, or substitution of at least one amino acid, a truncation, an extension, a chimeric fusion, a mutation, or polypeptides exhibiting post-translational modifications. Among the homologous polypeptides, those whose amino acid

sequences exhibit between at least (or at least about) 20.00% to 99.99% (inclusive) identity to the full length, native, or naturally occurring polypeptide are another aspect of the invention. The aforementioned range of percent identity is to be taken as including, and providing written description and support for, any fractional percentage, in intervals of 0.01%, between 20.00% and, up to, including 99.99%. These percentages are purely statistical and differences between two polypeptide sequences can be distributed randomly and over the entire sequence length. Thus, variant polypeptides can have 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent identity with the polypeptide sequences of the instant invention. In a preferred embodiment, a variant or modified polypeptide exhibits approximately 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to a natural polypeptide of the invention. Typically, the percent identity is calculated with reference to the full-length, native, and/or naturally occurring polypeptide (*e.g.*, those polypeptides set forth in SEQ ID NOs: 1, 3, or 4. In all instances, variant polypeptides retain at least one of the biological activities associated with thee polypeptides set forth in SEQ ID NOs: 1, 3, or 4. In some embodiments, variant polypeptides retain at least 2, and preferably all of the biological activities associated with the polypeptides set forth in SEQ ID NOs: 1, 3, or 4.

[0093] Variant polypeptides can also comprise one or more heterologous polypeptide sequences (*e.g.*, tags that facilitate purification of the polypeptides of the invention (see, for example, U.S. Patent No. 6,342,362, hereby incorporated by reference in its entirety; Altendorf *et al.* [1999-WWW, 2000] "Structure and Function of the F_o Complex of the ATP Synthase from *Escherichia Coli*," *J. of Experimental* Biology 203:19-28, The Co. of Biologists, Ltd., G.B.; Baneyx [1999] "Recombinant Protein Expression in *Escherichia coli*," Biotechnology 10:411-21, Elsevier Science Ltd.; Eihauer *et al.* [2001] "The FLAG™ Peptide, a Versatile Fusion Tag for the Purification of Recombinant Proteins," *J. Biochem Biophys Methods* 49:455-65; Jones *et al.* [1995] *J. Chromatography* 707:3-22; Jones *et al.* [1995] "Current Trends in Molecular Recognition and Bioseparation," *J. of Chromatography A.* 707:3-22, Elsevier Science B.V.; Margolin [2000] "Green Fluorescent Protein as a Reporter for Macromolecular Localization in Bacterial Cells," *Methods* 20:62-72, Academic Press; Puig *et al.* [2001] "The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification," *Methods* 24:218-29, Academic Press; Sassenfeld [1990] "Engineering Proteins for Purification," TibTech

8:88-93; Sheibani [1999] "Prokaryotic Gene Fusion Expression Systems and Their Use in Structural and Functional Studies of Proteins," *Prep. Biochem. & Biotechnol.* 29(1):77-90, Marcel Dekker, Inc.; Skerra *et al.* [1999] "Applications of a Peptide Ligand for Streptavidin: the *Strep*-tag", *Biomolecular Engineering* 16:79-86, Elsevier Science, B.V.; Smith [1998] "Cookbook for Eukaryotic Protein Expression: Yeast, Insect, and Plant Expression Systems," *The* Scientist 12(22):20; Smyth *et al.* [2000] "Eukaryotic Expression and Purification of Recombinant Extracellular Matrix Proteins Carrying the Strep II Tag", *Methods in Molecular Biology*, 139:49-57; Unger [1997] "Show Me the Money: Prokaryotic Expression Vectors and Purification Systems," *The Scientist* 11(17):20, each of which is hereby incorporated by reference in their entireties), or commercially available tags from vendors such as such as STRATAGENE (La Jolla, CA), NOVAGEN (Madison, WI), QIAGEN, Inc., (Valencia, CA), or InVitrogen (San Diego, CA).

[0094] The nomenclature used to describe peptide compounds follows the conventional practice wherein the amino group is presented to the left (the N-terminus) and the carboxyl group to the right (the C-terminus) of each amino acid residue. By way of example, amino acid substitutions can be carried out without resulting in a substantial modification of the biological activity of the corresponding modified polypeptides; for example, the replacement of leucine with valine or isoleucine, of aspartic acid with glutamic acid, of glutamine with asparagine, of arginine with lysine, and the like, the reverse substitutions can be performed without substantial modification of the biological activity of the polypeptides. The L-form of an amino acid residue is represented by a capital single letter or a capital first letter of a three-letter symbol, and the D-form, for those amino acids having D-forms, is represented by a lower case single letter or a lower case three letter symbol. Glycine has no asymmetric carbon atom and is simply referred to as "Gly" or G. Symbols for the amino acids are as follows: (Single Letter Symbol; Three Letter Symbol Amino Acid) A; Ala; Alanine: C; Cys; Cysteine: D; Asp; Aspartic Acid: E; Glu; Glutamic Acid: F; Phe; Phenylalanine: G; Gly; Glycine: H; His; Histidine: I; Ile; Isoleucine: K; Lys; Lysine: L; Leu; Leucine: M; Met; Methionine: N; Asn; Asparagine: P; Pro; Proline: Q; Gln; Glutamine: R; Arg; Arginine: S; Ser; Serine: T; Thr; Threonine: V; Val; Valine: W; Trp; Tryptophan: Y; Tyr; Tyrosine. Amino acid "chemical characteristics" are defined as: Aromatic (F, W, Y); Aliphatic-hydrophobic (L, I, V, M); Small polar (S, T, C); Large polar (Q, N); Acidic (D, E); Basic (R, H, K); Non-polar: Proline; Alanine; and Glycine.

[0095] In order to extend the life of the polypeptides according to the invention, it may be advantageous to use non-natural amino acids, for example in the D-form, or alternatively amino acid analogs, for example sulfur-containing forms of amino acids in the production of "variant polypeptides".

[0096] The subject invention also provides methods for bioremediation or phytoremediation of sites contaminated with metals, such as Zn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Hg^{2+} and Cu^{2+}, but not Ca^{2+}, Mg^{2+} and Mn^{2+}. Such methods comprise: a) identifying a site suitable for bioremediation; b) providing microbes or plants (including, but not limited to trees) that have been transformed to express a polypeptide according to the subject invention (*e.g.*, a multimeric construct as provided herein or a polypeptide according to SEQ ID NOs: 1 or 4); growing said microbes, plants, or trees at said site under conditions that allow for the accumulation and/or sequestration of metals that contaminate said site; and, optionally, harvesting said microbes, plants, or said trees to remove the metal contaminants from the site. Where the plants have been transformed with leaf specific promoters, and the metal accumulation occurs in the leaves of the plants, removal of metal contaminants is accomplished by collection of leaves that fall from the transformed plants. Where trees have been transformed with promoters specific to xylem, cells destined to function in the trunk of the tree, metal accumulates within the tree trunk over long periods of time. In this aspect of the invention, xylem tracheids and vessels act as minute capillary cation exchange columns (or cation sinks) and remove toxic metals from the xylem stream. In this case, metals are removed from the soil or water and are stabilized within the plant tissue (*e.g.*, the tree trunk) until the trees are harvested and the metals recovered.

[0097] Where the bioremediation and/or phytoremediation of aquatic sites or environments (*e.g.*, bodies of water (such as canals, rivers, lakes, streams, creeks, marine estuaries, bays, harbors, or other marine environments) swamps, bogs, and the like) is contemplated, transformed algae, including marine algae, can be utilized in the methods of the subject invention. Additionally, it may be desirable to utilize combinations of transformed microbes, algae, and higher plants (*e.g.*, trees that grow in an aquatic environment) in methods of bioremediation and/or phytoremediation as is taught herein in order to maximize the remediation process.

[0098] The subject invention also provides methods for targeting polypeptides to the cell wall of a plant cell. In this aspect of the invention, plants, trees, or cells derived therefrom are transformed with a genetic construct comprising the polypeptide of SEQ ID NO:1 fused to a heterologous

polypeptide. The plants, trees, or cells, are then grown under conditions that allow for the expression of the genetic construct. Expression of the genetic construct allows for the targeting of a desired polypeptide to the cell wall of the plant, tree, or plant cell.

[0099] The subject invention also provides articles of manufacture comprising protein chips/arrays, microcantilevers, or biosensors. Biosensors are electronic devices which produce electronic signals as the result of biological interactions. These biosensors comprise a biological receptor linked to an electronic transducer in such a way that biochemical activity is converted into electrical activity. The electronic component of the biosensors measure voltage, amperage, wavelengths, temperature, or mass. See, for example, Lowe, C. R. (1984) Biosensors 1:3-16. Thus, the subject invention provides a new class of plant metal-binding proteins as biosensors for heavy metal detection as well as methods for the detection of such metals comprising contacting a biosensor/protein chip/protein array with a sample suspected of containing a heavy metal selected from the group consisting of Zn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Hg^{2+} and Cu^{2+} and detecting the binding of the heavy metal to the proteins arrayed on the biosensor, protein chip, or protein array. In some embodiments of the protein array is addressable and the proteins are arrayed on a solid support. Methods of producing biosensors, microcantilevers, or protein arrays according to the subject invention are known in the art (see, for example, U.S. Patent Nos. 6,523,392; 5,719,324; 5,853,576; 5,824,512; and 5,877,014; each of which is hereby incorporated by reference in its entirety.

[00100] In one aspect of the invention, microcantilevers are used as physical, chemical, and biological sensors. When molecular binding occurs preferentially on one side of a cantilever, it undergoes bending corresponding to the differential stress occurring on either side of the cantilever. Moreover, the fundamental resonance frequency of the cantilever changes depending on changes in its spring constant and effective mass. By appropriately modifying the cantilever surface, binding-induced changes can be monitored. AgNt84-6 has been shown by equilibrium dialysis and mass spectral analysis to bind multiple atoms of Zn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Hg^{2+} and Cu^{2+} (but not Ca^{2+}, Mg^{2+} and Mn^{2+}). Thus, it is possible to detect such contaminants in the environment or in a biological sample.

[00101] There is also a need for a way to reliably label the endoplasmic reticulum as well as other organelles for electron microscopic (EM) study of cells. In the preparation of cells for EM analysis, a choice must be made between whether to fix cells to be studied so that membrane

structure is maintained, or to fix the cells so that the cellular proteins retain their ability to bind to antibody and thus can be analyzed by cytoimmunochemical methods. When fixed for cytoimmunochemical analysis, the ER, as well at other structural detail, is not well preserved and therefore not accurately represented in photomicrographs. Because of its signal sequence AgNt84 is targeted to the ER of plant cells and can be retained there in the presence of the ER-retention signal HDEL (C-terminal His-Asp-Glu-Leu amino acids) or KDEL (C-terminal Lys- Asp-Glu-Leu amino acids). AgNt84 has the potential to be useful in marking the ER for cytoimmunochemical analysis due to its metal-binding properties. Cells fixed for antibody staining could be incubated with low levels of metal chlorides, washed, and then analyzed by EM. The metal bound to AgNt84 in the ER would be opaque to the electron beam and thus provide a marker for ER. Likewise, AgNt84 could be modified to be targeted to any other location in the cell and serve as a histochemical marker for that location. Thus, in certain embodiments, the subject invention provides methods of visualizing the endoplasmic reticulum of at least one cell comprising the steps of: a) expressing a polypeptide comprising AgNt84 (SEQ ID NO: 1) or metal binding fragments thereof, multimers of AgNt84 or metal binding fragments thereof, or fusion proteins comprising AgNt84 or metal binding fragments thereof, in at least one cell, provided that said polypeptide contains the ER-retention signal HDEL or KDEL; b) fixing said at least one cell; c) optionally, contacting said at least one cell with an composition comprising a carrier and a metal ion; and d) visualizing said endoplasmic reticulum. In such methods, at least one cell is immunochemically stained with antibodies before or after said fixing step and the cells can be contacted with a composition comprising at least one metal ion and said visualization step is performed via electron microscopy. In some embodiments, the polypeptide is an AgNt84 fusion protein (comprising, for example, AgNt84 and a fluorescent polypeptide (e.g., fluorescent green protein) and the cells are not contacted with a composition comprising a metal ion. Thus, AgNt84, variants of AgNt84, fragments of AgNt84, fusion proteins comprising AgNt84, multimers of AgNt84, or metal binding fragments thereof can be used as a histochemical reagent for specific cells within tissues or to identify subcellular structures for analysis by Light Microscopy, Electron Microscopy, Energy Dispersive Spectroscopy (EDS) or Electron Energy Loss Spectroscopy (EELS).

[00102] AgNt84 also has the potential for use as a histochemical reagent for specific cells within tissues or to identify subcellular structures for analysis by Energy Dispersive Spectroscopy (EDS) and/or Electron Energy Loss Spectroscopy (EELS) (Feng, J. *et al.* [2004], "A system for

acquiring simultaneous electron energy-loss and X-ray spectrum-images", *Journal of Microscopy,* 215:92-99) and Secondary Ion Mass Spectrometry (SIMS) (Sodhi, R.N. [2004] "Time-of-flight secondary ion mass spectrometry (TOF-SIMS):- versatility in chemical and imaging surface analysis", *The Analyst* 129:483-487). In such an aspect of the invention, AgNt84, variants thereof, or fragments thereof are linked to signal peptides known to target certain subcellular organelles or other compartments of a eukaryotic cell (plant or animal) and used as described herein. Signal peptides suitable for use in this aspect of the invention are well-known to those skilled in the art and/or can be identified using any number of commercially available software suites (e.g., PSORT I and PSORT II (world-wide web site psort.nibb.ac.jp) or SecretomeP, Signal P, TargetP, or ChloroP (world-wide web site cbs.dtu.dk/services/).

[00103] Example 1 - To test the prediction that AgNt84 is synthesized on the endoplasmic reticulum (ER) and is targeted from there to the plasma membrane for export to the cell wall, we have made a series of plasmid constructs based on Gateway vectors (Invitrogen [2003] "Gateway technology: a universal technology to clone DNA sequences for functional analysis and expression in multiple systems"; Curtis, M and U. Grossniklaus [2003] "A Gateway TM cloning vector set for high-throughput functional analysis of genes in plants", *Plant Physiology* 133: 462-469; which is hereby incorporated by reference in its entirety).

[00104] In order to be able to visualize the location of AgNt84, we have made plasmid constructs that result in Green Fluorescent Protein (GFP) being fused to the C-terminal end of AgNt84. The vectors, listed in the attached tables, include both inducible (by estrogen or heat shock) and constitutive (35S) promoters. Using transient, ectopic expression in onion epidermal cells bombarded with plasmid-coated tungsten particles, we have shown that AgNt84 is targeted to the endoplasmic reticulum of the expressing cells, that it is retained in the ER in the presence of the ER-retention signal HDEL (Gomord, V., E. Wee and L. Faye [1999] "Protein retention and localization in the endoplasmic reticulum and Golgi apparatus" *Biochimie* 81:607-618) and that in the absence of HDEL it is exported to the cell wall.

[00105] It should be understood that the embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application. All patents, patent applications, provisional applications, and publications referred to or cited

herein are incorporated by reference in their entirety to the extent they are not inconsistent with the explicit teachings of this specification.